

#16  
BD  
7-19-02RECEIVED  
MAR 14 2002  
TECH CENTER 1600/2900

ENTERED

1632

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/950,051DATE: 03/04/2002  
TIME: 15:03:28Input Set : A:\34950209.app  
Output Set : N:\CRF3\03042002\I950051.raw

3 <110> APPLICANT: YERAMIAN, EDOUARD  
 5 <120> TITLE OF INVENTION: GENES AND THE PHYSICS OF THE DNA DOUBLE HELIX.  
 6 FORMULATION OF A PHYSICS-BASED GENE IDENTIFICATION  
 7 (PBG1) METHOD: AB INITIO IDENTIFICATION OF GENES IN  
 8 EUKARYOTIC GENOMES  
 10 <130> FILE REFERENCE: 03495-0209-00000  
 12 <140> CURRENT APPLICATION NUMBER: 09/950,051  
 13 <141> CURRENT FILING DATE: 2001-09-12  
 15 <150> PRIOR APPLICATION NUMBER: 60/232,146  
 16 <151> PRIOR FILING DATE: 2000-09-13  
 18 <160> NUMBER OF SEQ ID NOS: 9  
 20 <170> SOFTWARE: PatentIn Ver. 2.1  
 22 <210> SEQ ID NO: 1  
 23 <211> LENGTH: 213  
 24 <212> TYPE: DNA  
 25 <213> ORGANISM: Plasmodium falciparum  
 27 <400> SEQUENCE: 1  
 28 atgtgcatcac atgttaacgtt taatttttat ttgtaagata atgatttttag tgcgttgaaa 60  
 29 gtttaaggatga agaaattgtt ttctaagaaa aataattttct cttttctgct cttagcaat 120  
 30 gattcaaat ctgtaacaaa aaagtacata gttgatttga cttactaga taatattata 180  
 31 gaatccgtaa gaaataaaaag aaatataaaa aga 213  
 34 <210> SEQ ID NO: 2  
 35 <211> LENGTH: 212  
 36 <212> TYPE: DNA  
 37 <213> ORGANISM: Plasmodium falciparum  
 39 <400> SEQUENCE: 2  
 40 agtatatttt ttgtaacatc aaatttttgc atcggttgag ctccccaggt gcgttgaaa 60  
 41 ttaaggatga agaaattgtt tctaagaaa ataatttctc cttttctgct cttagcaat 120  
 42 attcaaatc tgtaacaaa aagtacatag ttgatttgac cttactagat aatattatag 180  
 43 aatccgaac caatacaat ttgcttctg tg 212  
 46 <210> SEQ ID NO: 3  
 47 <211> LENGTH: 225  
 48 <212> TYPE: DNA  
 49 <213> ORGANISM: Plasmodium falciparum  
 51 <400> SEQUENCE: 3  
 52 atgtatagac gcatacatat tattacattt gtaacgatca atcttttttt cttattatcc 60  
 53 ctatcccaca gatatacatga tagcgtccag aattttttga aggaagaaaa aaataactct 120  
 54 gataagttac aagatgatat agatgaggat gaggaaaaat attttgacga ggaaatttta 180  
 55 aggggaagcca aaaaaaaaag tgaagaatat gataaagacg atgaa 225  
 58 <210> SEQ ID NO: 4  
 59 <211> LENGTH: 225  
 60 <212> TYPE: DNA  
 61 <213> ORGANISM: Plasmodium falciparum

RECEIVED  
MAR 14 2002  
TECH CENTER 1600/2900

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set : N:\CRF3\03042002\I950051.raw

```

63 <400> SEQUENCE: 4
64 aaagaaaaac gcatacatat tattacattt gtaacgatca atcttttttt cttattatcc 60
65 ctatccccaca gatatacatga tagcgctccag aatttcttga aggaagaaaa aaataactct 120
66 gataagttac aagatgatat agatgaggat gaggaaaaat attttgacga ggaattttta 180
67 aggggaagcca aaaaaaaaaa aacattcttt tcacacgatt ttggtt 225
70 <210> SEQ ID NO: 5
71 <211> LENGTH: 600
72 <212> TYPE: DNA
73 <213> ORGANISM: Homo sapiens
75 <400> SEQUENCE: 5
76 ataaacattc tttagtccac acatagataa ataaataagg aagcaaatag acacacagaa 60
77 gagcgggaca gctctctctc ccggggagaat ttcaattagt aagtgtgaaa ggaacaaggc 120
78 agggaggaga atctctcaaca gagccccaca gggcagctgc gggcgaggcg cccggagggg 180
79 caccagcact gccggggcaaa cgctctggga gagcggggac agctgccaaag ttctagacat 240
80 gaccaattac agaggggaaa gggcgccacg cgagggaatg gcccgggcgt gtctacacct 300
81 atgccccacg cacactgctc ctgtgggatt cctcccccaa caagatgccc actctgacca 360
82 cgaggaacac tcagcaagt ccacgtggag gggcattcta caaaacaccc aaccgggtca 420
83 ggtgcgtgag gccaaggaga gattgggcaa ccgtcacaaa ccagagaagg cgaggagagc 480
84 ttctagccaa cgcatgtgg ggtcctgagc aggacccacc ggaagtgtgt gcagctgct 540
85 aaagaccgtc ctggctgaga agaaacagag cagcgctgct ttctcagagc tgggaaccga 600
88 <210> SEQ ID NO: 6
89 <211> LENGTH: 1980
90 <212> TYPE: DNA
91 <213> ORGANISM: Homo sapiens
93 <400> SEQUENCE: 6
94 acctcgatct cagactctct gcttccaaaa ccatgagaca cggaattttt gttgtgtgac 60
95 cagccagttt gtggtactgt ttgtcatggc agccccagga aaagaatata ttacagcata 120
96 caaaocataa ctccacattat cttaacttag aacccaaaaa aacctctctc cctaagcttt 180
97 caatcacaga ggcacattat ctgttccagc agcctagaaa accaaggccc aggggagcca 240
98 cccgtaggca cccactcccc atagcctggc acacacacac ggcagagcca cccacaggca 300
99 cccactctct atagtccagc acacacacgg cagagccacc cgcaggccacc cactccccat 360
100 agcccgccac acacgtggac catgccaccc tccacgtgcg cctgggggagc aaagcagcac 420
101 agcctgaact gcacctcagc tcttctctct gactctaaaa cacgcacatg cgccccaggc 480
102 caattccaag ttttgtaaac tgagcaacag ctcttgggaa acaaaaaaac agctactgtt 540
103 tattctcctg gagctgctgt tacaccccaa caaggaaagg agggcttctt gagcctctgt 600
104 ttctggacaa atgcaccaag gaggagtata aaagccccac aaaccggagc acctcactca 660
105 ctgcgtcaac cactccctcc catctccccc agctcaaccc ccagcaacgc cagctccacc 720
106 atgtcgtctg gctccagcga cctgagctac agcagccgcg tctgctcttc tggttctgt 780
107 gactcttgct ccgactctct gcaggtggac gactgccacg agagctgctg cgagcccccc 840
108 tgctgcgccc ccagctgctg cgccccggcc ccctgctgta gctgtgtgtg caccaccagt 900
109 agccgtgtgt ccagccccctg ctgccccagt acctgcgagc ccagccccgt ccaatcaggc 960
110 tgacacagct ccagcagccc ctggtgctgc cagcagctta gctgccagct ggttgcgtgt 1020
111 gctctctccc cctgcagaca ggcctgctgc gtgcccgtct gctgcaagac tgtctgctgc 1080
112 aagcctgtgt cgtgtgtgct cgtctgctgt ggggattctt catgtgccca gcagtctagc 1140
113 tgccagtcag ctgtgctcac ctctctcccc tgccagcagg cctgctgtgt gccatctgc 1200
114 tgcaagcctg tctgctctgc gattctctct tctgtctgac agcagctctg cgtgtgagc 1260
115 tgtgtgtcca gccctctgct ccaggcggtc tgtgagccca gccctctcca atcaggctgc 1320
116 atcagctcct gcaagcctcc gtgctgccag cagtctagct gccagccgcg ttgctgcacc 1380
117 tctctctctt gccagcaggc ctgctgctgt cccgtctgct gcaagactgt ctgctgcaag 1440

```

## RAW SEQUENCER LISTING

DATE: 03/04/2002

PATENT APPLICATION: US/09/950,051

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set : N:\CRF3\03042002\I950051.raw

```

118 cctgtgtgct ctgaggatgc ctcttcacgc tgcagcagc ctagtgcgca gccggttgc 1500
119 tgcacctctc ctccctgcga gcaggcttgc tgtgtgcctg tctgtgcgaa gccgtgtgtgc 1560
120 tgcagacctg tgggtctctg gcccatctgc tctggggctt cctctctgtg ctgcgcagcag 1620
121 tctagtctgc agccagcttg ctgcacctcc tcccaaagcc agcagggtcg ctgctgtgcc 1680
122 gtctgtctga agctgtgag ctgtgtgcct gtttgcctg gggcttctc ttcattgtgc 1740
123 cagcaatcta gctgcagacc agctgtctgc accacctctc gtgcagacc cctctctc 1800
124 gtgtccctcc tctgccgcc cgtgtgcagg cccgctctc cgtgtccctg cctctctc 1860
125 tgtgtcccca cctctctgc ccaaccagc tgtgtccgc cagctctc cgtgtctc 1920
126 ctctgacgcc cgtgtgtgc cgcgccagc tgtgtgagc tccgtcagg tcagaagccc 1980
129 <210> SEQ ID NO: 7
130 <211> LENGTH: 1800
131 <212> TYPE: DNA
132 <213> ORGANISM: Homo sapiens
134 <400> SEQUENCE: 7
135 ggatgagagg gggactcatg gaggaacagc cagccttga ccttgagatg gccctgcagg 60
136 gagggttaact gaaaatttac ccactgggga cagtgccta ctactaaaa cagtctcagg 120
137 caccacgcga gccctgggaa ggccatcccc ccagaaaaac cccaggctct cagcaggggc 180
138 ttgtccactc gtgccctcca gtgtgcgcca tgtcaacctc acctaaagag ggcctgcagg 240
139 accgtctctg aggtgcggac tctgggtctc tgcagcccat cggaacctg gtgcccagc 300
140 aggagggtct ggggcaatgc cagttttggg gaattcatgt catccatcca cccactccat 360
141 gatgtttct tctgtatoga gtccctgtgc tccgcgcag gtgcagacac cctccctct 420
142 ccccccgcat tgcctgtgaa cgggcagaa cctcgggcgg gcggcacaca gggagggtga 480
143 ccaggccttg aggtctgagt gcccgagccc caggccagct tcttggaaag tgacctctc 540
144 ggggtgggctc tcccaggttg gacagtgggt gggacagtcc tggggcctgg agagcccac 600
145 agccccagggc agcgacgcca atgaccaggg tcaggaaagc acaggcatgg aggtgcaggc 660
146 gggactgagc cttcctgggc gtggctgtga gtccacctg gtgacccct ggaggagtta 720
147 gggcactgtc cccctgacct tctagtgtta gtacctatt catagaaaaa gtcattggcta 780
148 gagagcaatc ttagctcaaa acctgtatc cccaggagca ctacagaaaa agagaatcag 840
149 gcgaccaaag ggaagttatt ggggagcagg aggaggtgct gacaggttca agtcaggggc 900
150 aagtgcacct gggcagagaa gctgggaggg aggcacaggg acccaacagg caggtggggc 960
151 cctctgggga ggcagagagt ggggagcttc gaggatggag attcctggga gtatggagg 1020
152 ggggtgcacc tcagcacatg gggggccctg ccaagcggg ggcaacctcc taaccagat 1080
153 caggaccagt tggccctggg ggaatgtcac atagcaact ggaactcctg cctgagcaga 1140
154 ggcctcagca ggcaggcggg gagcacggg ggcggcagag gagggaacag caggaggccg 1200
155 ggcggcagca gctggccttg taggaggagg caggggcaca gcaggaggag atgggcaggc 1260
156 agcaggcggg cctgcataat gggcggcaga gggggacac ggaaggaggg ggtctgcagg 1320
157 aggaggtggt gcagcaagcc ggcgtgacag tagactgtcg gcagatgaa gtggaagccc 1380
158 cagagcagac gggcacacag cagatgggtt tgaagcagac aggtctgcga cagacaggca 1440
159 cgtagcagga ctgctggcag ggggaggagg tgacagcaag cggctggcag ctgaaatgct 1500
160 ggcagcatga agaggaatcc tcagaaacag tgggcacaca gcacacgggc ttgcagcaga 1560
161 caggcacaca cgtctggcag gcagactgc tggcaggagg aagagccaca cgaagtggc 1620
162 actgtcggca gcatgaagag gaactcttag agcaggtggg caggcagcac acaggtctg 1680
163 agcagcaggg cagcagcagc gcctgctggc agggggagga ggcgcagcaa gccgctggc 1740
164 agcagcaggg cgtgcagagg ctggtgcagc ctgattggca ggggctgggc tcacaggccg 1800
167 <210> SEQ ID NO: 8
168 <211> LENGTH: 1620
169 <212> TYPE: DNA
170 <213> ORGANISM: Plasmodium falciparum
172 <400> SEQUENCE: 8

```

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002  
 TIME: 15:03:28

Input Set : A:\34950209.app  
 Output Set : N:\CRF3\03042002\I950051.raw

```

173 tatgcacgtg tgtgtgtgta taatatataa ttctttatct catattacct tgtctacata 60
174 tgggacattt cgtattttgc atcttaacaa catttgcaca ctctcaaaaa aaaaaaataa 120
175 aaaaattatat atatatatat atatatata tatcacatga ttacttttta atattccat 180
176 tatttctgtt taaagtaacca cattacaagg acacatatgt ctacaaggta aaattgctgt 240
177 atcttttttc tccgtcaggc atattacaca ttcccttaaa aataaaacat aataataat 300
178 aagcatalaa atataatagt atataaatg tacattcata cgtataaaatg aataaaaaa 360
179 tacatatata tatatatata tatatatata tgaatgacac acctgtcaga caataaata 420
180 tttaacaggat cggtttttgt ttctggagat ttctctatct caaaaatttc ttgaacctca 480
181 aatgatctgt taccaaatg aatcttctgt ctataaagaa ttattttata ttattcttga 540
182 ttgtttttat tctctgtcat ctgattttct tgtaagttag catagtgtga ttgagccctga 600
183 agtatgggag taccaatcgc tttaaggact attaaaatag gaatgatata ttcataacta 660
184 ggttttacatt ttatattcat tatattcatg ttattcaaat ttattccttc atttaattga 720
185 ctcatataaa ctgtgtttaa attcttcgga aatatitttg taactgtttg atacttttga 780
186 ggacaatatt tagaagaata agaaaattaa caaaacgaac ataaaaaat tatttaaaaa 840
187 aaaaatataaa aaaaaaaa aaattgataaa taacacataaa attctacat aataaatcta 900
188 atatatatta atatatatat atatatatat atatatatat tataatacat ttctttttat 960
189 tttaaaaaata ttacatggtt tcttttgctt ctgttaactc ctctttacaa caaaaataaa 1020
190 tagatccctc tacatcatgt aaggcatcaa aaatgaaatt tattaaataa tattatttc 1080
191 ccaactttat aatttttaat gtctttctac gtaagtttaag tgaattttct acaacggtg 1140
192 ttctttgaat atttatattt ggtccttgcc atattagttc acacaagtgt ggtgtattgt 1200
193 ttctctatca taagtcaaaa atagaagaga aaaaaatga gaagtataa gaagtataa 1260
194 tacacataaa tgtacatata tatatatata tatttatatt ttatatatt ttgcttctta 1320
195 taaagtgtat acatttatta tgggattatc aaatgatatt gtttgggct ttttggaac 1380
196 aaaaaaaa aaaaaaaa aaagttttt aataaaatat ggtttatata gataataatg 1440
197 agaagatata tatatatata tatatggagc aaaaatttat cattcttttt ttctttttaa 1500
198 aaaccttaca ccatcatcat aataatcact tcgattgtta gataacgatg aacctattat 1560
199 tcccattata caacaagaaa attataaaat gtaacatata tatttttaag cattataaat 1620
202 <210> SEQ ID NO: 9
203 <211> LENGTH: 1560
204 <212> TYPE: DNA
205 <213> ORGANISM: Plasmodium falciparum
207 <400> SEQUENCE: 9
208 actataaat aggttgatata caaatatata ttatttatat ttattttgt taataaatc 60
209 ttggttaacaa ttctttcca tcatcatagg gaagatattt actcttatct ttccggacgt 120
210 agcctgaat atttaaaat aataaaataa atatatatat atataatata tataattacc 180
211 ttgttgatatt atatatatat atatatatat attttttttt tcatttacct ccatgatga 240
212 ttgtctgaca caatgctggt aaacgacgat ttgatttttt ctggtttataa aaatcagccc 300
213 tatataaaa aagcaaaaata taaaacgtat atatatatat atatatatta tatattttat 360
214 ttgtgtgta caagaattac catataaaat aggtctgttaa aaaaatgatca ctgtgaaat 420
215 taaatatatg tctcatttga agaatactat atttacttaa ctctgttaca attttccccc 480
216 attgctctac tttttttcca tataaatgcc tgaaaagggg aaaaatacaa aataaattaa 540
217 aataataata tatatatata tatatatatta ttctttttt attttttttt tttaagttaa 600
218 actaacctcc ataatatata ctcaacaatt gatggaaaaa tatctacatt gttagccgct 660
219 tegtctaatg atacaaaaaa ctgaaaaaag aaaaataaac attcaagtgt attatcatta 720
220 ttgtttattac atatatgtat ataaataaat atatatatat atatatatat tatattatat 780
221 atataataac aaatgatata ttctttttta ttaaaaaaaa aaaaagctc actccaaagt 840
222 catattcttg acaattcaaa agttctgcac taaaagaata ttccaggaca tcttttaaaa 900
223 gcatccaatc acgaacaatt ctccagctta tatcaaaat atatgtattt ataatactc 960
224 ctgtgatact ttcatgtctt aatcttttat taattatcca aagatgcaaa acataaaat 1020

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw

```
225 acattttcat attaaaagat tcattttatc gaaataagtt taatatctct tcattttcta 1080
226 atctttcaat tattaaatgt actaatctcc atgcacatcc tccatattct gtatgttttc 1140
227 taaaaaata ttaaaatata tatattttta ttttatatat aaacataaat tcaaaatgaa 1200
228 acttttttaa catattagac gtacatatgc tatgtttttt aaaatattat atatatatat 1260
229 tatgaagaat tgttttttta caaaaaaaat ttatatatca tagaacttat tatagaacga 1320
230 tcacattgtg gtatagggtat tgcatatott gaagattctc taataactaat ttttgatttc 1380
231 ttatctcat acaaactttt actttcgata atttttaaat tatcatatac ttttatttca 1440
232 ctagaaaaat agcgatagtt aatatatott ctctttaaaa caaattttgt acattttatac 1500
233 cacatgacaa caaaatttaa ttaaaaatta atgaaaaaaa aaaaaaaaaa aaaaaaaaaa 1560
```

VERIFICATION SUMMARY

DATE: 03/04/2002

PATENT APPLICATION: US/09/950,051

TIME: 15:03:29

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw